© 1993 Oxford University Press

*Nucleic Acids Research, 1993, Vol. 21, No. 17* **4085–4092**

# Purification and cDNA cloning of HeLa cell p54$^{nrb}$, a nuclear protein with two RNA recognition motifs and extensive homology to human splicing factor PSF and *Drosophila* NONA/BJ6

Benhao Dong, David S.Horowitz, Ryuji Kobayashi and Adrian R.Krainer*
Cold Spring Harbor Laboratory, PO Box 100, 1 Bungtown Road, Cold Spring Harbor, NY 11724-2208, USA

Received April 27, 1993; Revised and Accepted June 19, 1993

GenBank accession no. L14599

## ABSTRACT

**While searching for a human homolog of the *S.cerevisiae* splicing factor PRP18, we found a polypeptide that reacted strongly with antibodies against PRP18. We purified this polypeptide from HeLa cells using a Western blot assay, and named it p54$^{nrb}$ (for nuclear RNA-binding protein, 54 kDa). cDNAs encoding p54$^{nrb}$ were cloned with probes derived from partial sequence of the purified protein. These cDNAs have identical coding sequences but differ as a result of alternative splicing in the 5′ untranslated region. The cDNAs encode a 471 aa polypeptide that contains two RNA recognition motifs (RRMs). Human p54$^{nrb}$ has no homology to yeast PRP18, except for a common epitope, but is instead 71% identical to human splicing factor PSF within a 320 aa region that includes both RRMs. In addition, both p54$^{nrb}$ and PSF are rich in Pro and Gln residues outside the main homology region. The *Drosophila* puff-specific protein BJ6, one of three products encoded by the alternatively spliced *no-on-transient* A gene (*nonA*), which is required for normal vision and courtship song, is 42% identical to p54$^{nrb}$ in the same 320 aa region. The striking homology between p54$^{nrb}$, PSF, and NONA/BJ6 defines a novel phylogenetically conserved protein segment, termed DBHS domain (for *Drosophila* behavior, human splicing), which may be involved in regulating diverse pathways at the level of pre-mRNA splicing.**

## INTRODUCTION

Pre-mRNA introns are present in a small proportion of *S.cerevisiae* genes, almost always as single short sequences near the 5′ end of the gene. The 5′ and 3′ splice sites and the branch site are very highly conserved in budding yeast. In metazoans, intronless genes are infrequent, whereas genes with multiple introns are common, and there is enormous variation in intron sizes. The conserved elements are highly degenerate, and complex patterns of alternative splicing are common. In spite of these substantial differences, the metazoan and yeast pre-mRNA splicing pathways appear very similar (reviewed in 1,2). The structures and functions of the spliceosomal snRNAs are well conserved from yeast to man, and thus, e.g., human U2 snRNA can functionally replace its *S.cerevisiae* counterpart *in vivo* (3). There have also been reports of a putative human homolog of the yeast PRP8 protein (4–6), of yeast homologs of the 70K and A polypeptides of metazoan U1 snRNP (7,8) and the D1 core snRNP polypeptide (9), and of a possible yeast homolog of the human U2AF$^{65}$ splicing factor (10).

Genetic approaches have facilitated the identification of more than 20 *PRP* genes required for pre-mRNA splicing in *S.cerevisiae* (reviewed in 11). One of these genes, *PRP18*, encodes a 28 kDa U5 snRNP-associated protein involved in the second step reaction of pre-mRNA splicing *in vivo* and *in vitro* (12–14). In the absence of PRP18 protein, splicing intermediates accumulate, and upon addition of recombinant PRP18 protein expressed in bacteria, the second step reaction is restored (15). If there is substantial phylogenetic conservation of protein splicing factors, it may be possible to take advantage of the *S.cerevisiae prp* mutant strains, genes, proteins or antibodies to identify their metazoan homologs.

When antibodies against PRP18 (14) were used to probe HeLa cell extracts on Western blots, a band of apparent molecular weight 60 kDa was consistently observed. The same antibodies also produced a nucleoplasmic speckled pattern upon indirect immunofluorescence staining of HeLa cells, typical of the pattern generated with antibodies against nucleoplasmic snRNP particles and other splicing factors (reviewed in 16). We report the purification of this cross-reactive protein, termed p54$^{nrb}$, and the cloning of the corresponding HeLa cell cDNAs. Surprisingly, the protein sequence of p54$^{nrb}$ bears no sequence homology to PRP18. It is, however, highly homologous to the recently identified human splicing factor PSF (17). The p54$^{nrb}$ sequence

* To whom correspondence should be addressed

contains two RNA recognition motifs (RRMs, reviewed in 18), strongly suggesting that it is an RNA-binding protein.

## MATERIALS AND METHODS

### Antibodies

The preparation of rabbit antisera against *S.cerevisiae* PRP18 protein expressed in *E.coli* has been described (14). Anti-PRP18 IgGs, which recognize yeast PRP18 by Western blotting, were purified from rabbit sera by Protein A-Sepharose chromatography.

### Western blotting

Protein samples were separated by SDS-PAGE and electroblotted onto nitrocellulose membranes (Schleicher & Schuell) in a semi-dry apparatus (Owl Scientific). The filters were probed with anti-PRP18 antibody at 1:500 dilution, using the ProtoBlot AP system (Promega) with 5% nonfat milk as the blocking agent.

### Purification of $p54^{nrb}$

HeLa cell nuclear extract was prepared essentially as described (19). The crude nuclear extract was dialyzed against 100 volumes of buffer 1 (20 mM Hepes-KOH, pH 8 + 0.1 M KCl + 0.2 mM EDTA + 20% glycerol + 1 mM DTT + 0.2 mM PMSF) for 12 hours at 4°C. The dialysate was centrifuged at 16,000×g for 30 min. The resulting precipitate was resuspended in buffer 1 containing 6 M urea and 100 mM NaCl, and loaded at 0.25 ml/min onto a Mono Q column (Pharmacia) pre-equilibrated in the same buffer. The flowthrough fractions were pooled, combined with an equal volume of buffer 1 + 6 M urea to adjust the salt concentration to 50 mM, then loaded onto a Mono S column (Pharmacia). Proteins bound to the column were eluted by a 50 mM to 1 M salt gradient in buffer 1 + 6 M urea, at a flow rate of 0.25 ml/min. $p54^{nrb}$ eluted at around 170 mM NaCl. The Mono S fraction containing $p54^{nrb}$ was purified further by preparative SDS-PAGE (20). The gel was stained briefly in 0.25% Coomassie Blue R250 in deionized water; $p54^{nrb}$ was excised, crushed in a Dounce homogenizer, and recovered from the gel as described (21).

### Peptide sequencing of $p54^{nrb}$

12 μg of purified $p54^{nrb}$ was treated with 250 ng of *Achromobacter lyticus* protease I (lysyl-endopeptidase; Wako) in 0.1% SDS + 0.15 M Tris HCl, pH 9, at 30°C for 18 hours. In some cases the digestion was performed in an excised gel slice after staining with Coomassie Blue G (Aldrich), essentially as described (22). The separation and sequence analysis of peptides were performed at the Cold Spring Harbor Protein Chemistry Facility. The peptides were separated by reverse-phase HPLC in a Hewlett-Packard 1090 system using a DEAE pre-column (Brownlee GAX-013, 3.2×15 mm) and a C18 column (Vydac, 5 μm, 300 Å, 2.1×250 mm) (23). The peptides were eluted with an acetonitrile gradient and monitored at 214 nm. Amino acid sequence analysis of selected peaks was performed by automated Edman degradation using an Applied Biosystems model 470A sequencer with an on-line 120A PTH analyzer.

### Oligonucleotide primers

Oligonucleotides were synthesized by S.Teplin at the Cold Spring Harbor Oligonucleotide Facility. P54BAM contains three extra nt and a *Bam*H I recognition site 5′ to the degenerate coding sequence derived from peptide 1. P54HIN contains three extra nt and a *Hind* III site 5′ of the degenerate antisense sequence derived from peptide 3. P54PRMNDE consists of the first 18 nt of the $p54^{nrb}$ ORF, preceded by an *Nde* I recognition site and three extra nt. P54PRMBAM has the sequence complementary to the last 13 nt of the $p54^{nrb}$ ORF including the stop codon, plus a *Bam*H I site and three extra nt at the 5′ end of the primer. The sequences of these oligonucleotides are as follows (r=a+g; y=t+c; m=a+c; n=t+g; x=a+c+g+t):

| | |
|---|---|
| P54BAM: | cttggatccccxmgxttygcxcarcc |
| P54HIN: | ctcaagcttcnrtcxacytgrtcytg |
| P54OLI41: | gccttccagcgcatggcatattcatactcaaaggagccagg |
| P54PRMNDE: | gtccatatgcagagtaataaaact |
| P54PRMBAM: | gctggatccttagtatcggcga |

### PCR isolation of a specific $p54^{nrb}$ probe

8 μg of HeLa polyadenylated RNA was annealed to 5 pmol of oligonucleotide P54HIN by heating in 10 μl of 10 mM Tris, pH 7.6 + 150 mM KCl at 85°C for 2 min, 65°C for 5 min, then gradually cooling the mixture to 40°C. The mixture was then incubated with 5 units of AMV reverse transcriptase (Life Sciences) at 42°C for 2 hours in 50 μl of 50 mM Tris, pH 8.3 + 120 mM KCl + 8 mM $MgCl_2$ + 1 mM dNTPs + 0.1 mg/ml BSA + 1 mM DTT. The mixture was treated with NaOH to degrade the RNA, neutralized, phenol extracted and ethanol precipitated. The cDNA was resuspended in 100 μl water.

1 μl of the first strand cDNA was mixed with 100 pmol each of oligonucleotides P54BAM and P54HIN, in 100 μl of 20mM Tris, pH 8.3 + 0.2 mM dNTPs + 2 mM $MgCl_2$ + 0.1 mg/ml BSA + 5 units of AmpliTaq polymerase (Perkin Elmer Cetus). The mixture was overlaid with 100 μl of light mineral oil and incubated in a Hybaid OmniGene thermal cycler. Cycling times for denaturation, annealing, and extension were, respectively, 1.5 min, 5 min and 5 min for the first cycle, and 0.5 min, 4 min and 4 min for 40 repetitive cycles. Final extension was for 20 min. Denaturation was at 95°C, annealing at 50°C or 55°C, and extension at 72°C.

A discrete PCR DNA fragment of approximately 118 bp was purified on a 1.8% agarose gel, digested with *Bam*H I and *Hind* III, and subcloned into the corresponding sites of pGEMEX (Promega) to generate pGEMEX-sub1. Nucleotide sequencing of the insert was performed with a Sequenase II kit (USBC) using pGEMEX-sub1 as a double-stranded template. SP6 and T3 promoter primers were used to sequence both DNA strands. Unambiguous nucleotide sequence corresponding to the region between peptides 1 and 3 was obtained, part of which was used to synthesize an antisense 41-mer oligonucleotide, P54OLI41.

### Library screening

An amplified oligo(dT)-primed HeLa cell cDNA library subcloned with *Eco*R I linkers into λ1149 was kindly provided by C.Schneider. Recombinant phage were plated at a density of $10^5$ pfu per 15 cm Petri dish in *E.coli* LE392. Plaque lifts were prepared and probed on duplicate filters (GeneScreen, NEN) as described (24). The probe was 5′ end-labeled oligonucleotide P54OLI41. Five positive plaques were isolated and purified, and their DNA was extracted by standard methods (25).

### Nucleotide sequencing of cDNA clones

DNA purified from recombinant phage isolates was digested with *Eco*R I, and the inserts were subcloned in both orientations into the *Eco*R I site of pGEMEX to generate pGEMEX-$54/1^+$,

pGEMEX-54/1⁻, pGEMEX-54/2⁺, pGEMEX-54/2⁻, pGEMEX-54/3⁺, pGEMEX-54/3⁻, pGEMEX-54/4⁺, pGEMEX-54/4⁻, pGEMEX-54/5⁺, and pGEMEX-54/5⁻ (the minus sign indicates that the sense strand is in the same orientation as that of T7 gene 10). An Erase-A-Base System kit (Promega) was used to generate a nested set of exonuclease III deletions in the insert DNA of pGEMEX-54/4⁺ and pGEMEX-54/4⁻. A series of pGEMEX-based plasmids containing variably shortened insert DNA was obtained and prepared as double-stranded DNA templates by alkaline lysis and PEG/NaCl precipitation (25). The overlapping sequences obtained were combined to give complete sequence information for the insert DNA of pGEMEX-54/4. Since restriction analysis of the other four inserts indicated that they were all highly related to that of pGEMEX-54/4, they were sequenced from both ends until the sequence overlapped with that of the pGEMEX-54/4 insert.

### Expression of $p54^{nrb}$ in bacteria

1 ng of pGEMEX54/4⁺ was used as template, and mixed with 100 pmol each of oligonucleotides P54PRMNDE and P54PRMBAM in a PCR reaction with Vent polymerase (New England Biolabs) to amplify the open reading frame of $p54^{nrb}$. The PCR product was digested with *Bam*H I and *Nde* I, and inserted into the corresponding sites of the expression vector pET-9c (Novagene). The resulting plasmid pET-9cp54 was transformed into *E.coli* BL21(DE3)-pLysS cells (26). Transformants resistant to both kanamycin and chloramphenicol were selected. Bacterial cultures grown in M9ZB (26) were induced for 3 hr with 0.4 mM IPTG when their $O.D_{600}$ reached 0.6. IPTG-induced and mock-induced cells were boiled for 5 min in SDS-PAGE sample buffer (20) and analyzed by Western blotting.

### Northern blotting

Polyadenylated RNA was obtained from HeLa cells grown in suspension culture using RNAgents and PolyATract kits (Promega). 5 μg of polyadenylated RNA was fractionated by agarose-formaldehyde gel electrophoresis. After ethidium bromide staining to visualize the markers, the RNA was transferred onto a GeneScreen Plus membrane (NEN) in a vacuum blotting apparatus (Pharmacia) in 20× SSC at 50−55 cm $H_2O$. The hybridization probe was made by digesting pGEMEX-54/4⁺ with *Sac* I, which cuts once in the cDNA at position +1021 (Fig. 3), and once in the polylinker, downstream of the insert. This fragment was labeled with α-$^{32}$P-dATP by random priming using a Prime-a-Gene labeling system (Promega). Aqueous hybridization and washing were carried out as described (25).

### Southern blotting

20 μg aliquots of human sperm DNA were separately digested with *Eco*R I, *Hind* III, *Nco* I, and *Pst* I. The digested DNA was separated by agarose gel electrophoresis and transferred onto a GeneScreen Plus membrane (NEN) in a vacuum blotting apparatus (Pharmacia) as described by the manufacturers. The blot was pre-hybridized with 6× SSC + 0.5% SDS + 5× Denhardt's solution + 100 μg/ml sonicated denatured salmon sperm DNA at 68°C for 2 hours, then hybridized with the same hybridization probe used for Northern blotting, in 6× SSC + 0.5% SDS + and 100 μg/ml salmon sperm DNA at 68°C for 16−18 hours. The membrane was removed from the hybridization solution and washed with two changes of 100 ml of 2× SSC at room temperature for 5 min each, then washed twice with 2× SSC + 0.1% SDS at 70°C for 30 min each, and finally washed twice with 0.1× SSC at room temperature for 30 min each.

## RESULTS

### Identification and purification of human $p54^{nrb}$

The product of the *PRP18* gene of *S.cerevisiae* is involved in the second cleavage-ligation reaction of pre-mRNA splicing *in vivo* and *in vitro* (12−15). PRP18 protein has been expressed in *E.coli* and the recombinant protein was used to generate polyclonal antibodies (14). Splicing intermediates accumulate in splicing reactions with extracts immunodepleted of PRP18, and can be chased to spliced products upon addition of purified recombinant PRP18 (15).

To determine whether antibodies to yeast PRP18 recognize human proteins, HeLa cells were analyzed by Western blotting (Fig. 1A). A prominent band with a relative electrophoretic mobility of approximately 60 kDa was detected. Although several weaker bands were also detected, their immunoreactivity varied depending upon the sample, blocking conditions, and antiserum. In contrast, the above-mentioned polypeptide, hereafter referred to as $p54^{nrb}$, was the only band recognized by two different antisera, and was reproducibly observed, even when affinity-purified antibodies were used (data not shown). When HeLa cell nuclear extract and a cytoplasmic S100 fraction (19) were compared, $p54^{nrb}$ was found to be enriched in the nuclear fraction (Fig. 1B).

Using relative electrophoretic mobility and immunoreactivity with anti-PRP18 antibody as fractionation criteria, we purified $p54^{nrb}$ from HeLa cell nuclear extracts. After dialysis in low salt, the crude nuclear extract was centrifuged to separate soluble and insoluble proteins. The precipitate was found by Western blot analysis to be enriched in $p54^{nrb}$ (Fig. 1C). The pellet was resuspended in buffer 1 containing 6 M urea to promote solubility and to disrupt protein−protein and protein−nucleic acid interactions, and thus improve chromatographic resolution. The

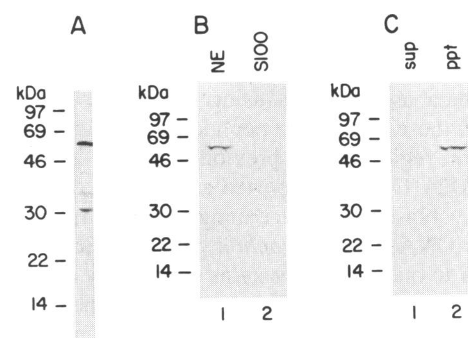


**Figure 1.** A nuclear protein from HeLa cells ($p54^{nrb}$) cross-reacts with antibodies to yeast PRP18. Protein samples were fractionated by SDS-PAGE, and analyzed by Western blotting with IgGs purified from anti-PRP18 serum. **(A)** $4\times10^4$ HeLa cells boiled in SDS sample loading buffer. **(B)** lane 1: HeLa cell nuclear extract; lane 2: HeLa cell cytoplasmic S100 extract ($6\times10^5$ cell equivalents each). **(C)** Partial purification of $p54^{nrb}$ by precipitation at low ionic strength. Nuclear extract was dialyzed against low salt buffer and the resulting insoluble material was recovered by centrifugation. Lane 1: supernatant; lane 2: precipitate ($2\times10^6$ cell equivalents each).

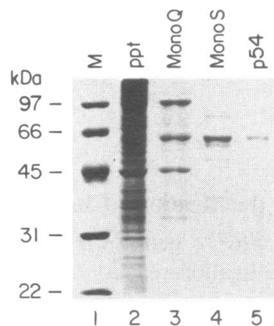


**Figure 2.** Purification of HeLa p54$^{nrb}$. HeLa cell nuclear extract was fractionated and p54$^{nrb}$ followed by Western blot analysis. Sequential peak fractions analyzed by SDS-PAGE and Coomassie Blue R250 staining are shown. Lane 1: protein size markers; lane 2: low salt precipitate ($2\times10^6$ cell equivalents); lane 3: flowthrough fraction from Mono Q column ($1.3\times10^7$ cell equivalents); lane 4: Mono S peak gradient fraction ($2\times10^7$ cell equivalents); lane 5: gel-purified p54$^{nrb}$ ($3.5\times10^6$ cell equivalents).

denatured material was purified by two ion exchange chromatography steps in the presence of 6 M urea (Fig. 2). p54$^{nrb}$ was recovered in the Mono Q flowthrough (lane 3), and is one of the prominent polypeptides in this fraction. The Mono Q fraction was loaded onto a Mono S column and p54$^{nrb}$ eluted at approximately 170 mM salt (lane 4). Finally, p54$^{nrb}$ was separated from minor contaminants in the Mono S fraction by preparative SDS-PAGE (lane 5).

## Cloning and analysis of p54$^{nrb}$ cDNAs

Purified HeLa p54$^{nrb}$ was found to have a blocked N-terminus (data not shown). To obtain internal amino acid sequence the purified protein was digested with lysyl-endopeptidase and the resulting peptide mixture was fractionated by reverse-phase HPLC. Sequencing of selected peak fractions yielded the following peptides (bracketed Ks indicate the expected Lys residues preceding the cleavage site; X denotes ambiguous sequence):

peptide 1: (K)XXEQPPRFAQPGSFEYEY
peptide 2: (K)ALIEMEK
peptide 3: (K)QQQDQVDRNI

Protein database searches using the BLAST program (27) showed that these three short peptides were highly homologous to a common region in two previously cloned cDNAs. One of these, HUM241D5, was reported as a partial cDNA encoding a human myoblast cell-surface antigen (28). The other, known as BJ6 or NONA, is a *Drosophila* puff-specific protein, which corresponds to one of three proteins encoded by the alternatively spliced transcripts of the *no-on-transient A* gene (29,30). From the alignment of the three p54$^{nrb}$ peptides with these two proteins, the expected order of the peptides in the human p54$^{nrb}$ sequence is 1−2−3 (see below).

To clone human cDNAs encoding p54$^{nrb}$ we designed two degenerate oligonucleotide pools: P54BAM (sense) and P54HIN (antisense), corresponding to peptides 1 and 3, respectively (see Materials and Methods). First, P54HIN was used as the primer to synthesize p54$^{nrb}$ first strand cDNA from HeLa poly-adenylated RNA by reverse transcription. Both oligonucleotides were then used in a polymerase chain reaction to amplify the first strand cDNA. Analysis of subcloned PCR products yielded

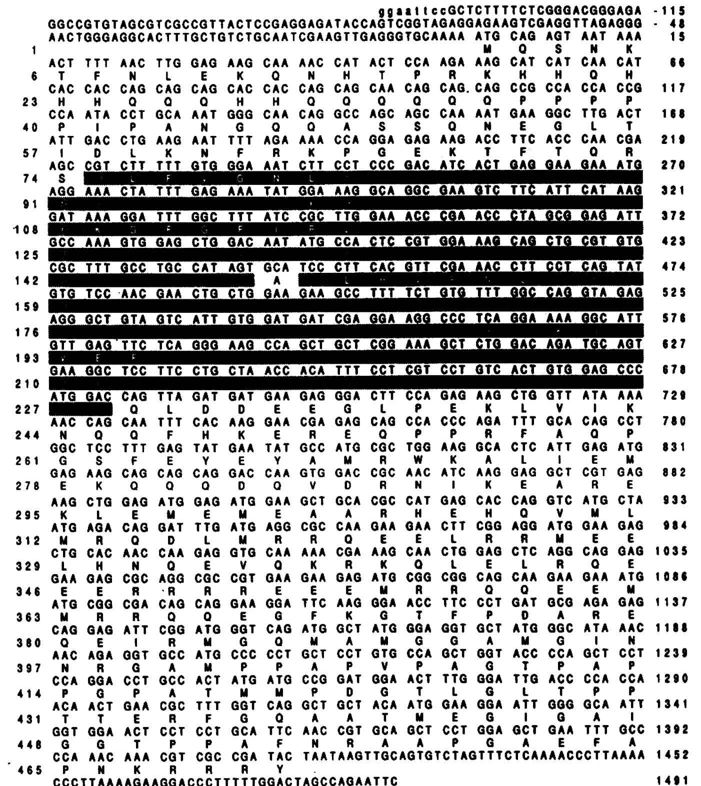

**Figure 3.** Nucleotide sequence and translation of p54$^{nrb}$ HeLa cDNA clone 4. The nucleotide sequence of the sense strand is shown, with the translated protein sequence in the one-letter code below the corresponding codons. Nucleotide residues are numbered on the right, and amino acid residues on the left. The two RRMs, separated by a single Ala residue, are highlighted with gray background. The RNP-2 and RNP-1 submotifs of each RRM are highlighted with black background. Except for the 5′ untranslated region, the five cDNA clones share the same sequence. The *Eco*R I linker at the 5′ end is shown in lowercase.

unique nucleotide sequence of the region between the degenerate oligonucleotides. A 41-mer antisense oligonucleotide, P54OLI41, was synthesized according to this unique sequence, and was end-labeled as a probe to screen a HeLa cDNA library in bacteriophage λ. Dozens of positive plaques were observed among $6\times10^5$ plaques screened on duplicate filters, suggesting that p54$^{nrb}$ mRNAs are abundant. Five positive clones picked at random were plaque-purified, and the insert DNAs were subcloned into plasmid pGEMEX. Restriction analysis indicated that the five inserts were highly related (data not shown).

Nucleotide sequencing of the cDNA inserts showed that the five clones represent three independent cDNAs that share a common open reading frame (ORF) and a short 3′ tail (Fig. 3). The three classes of cDNA clones differ in their 5′ untranslated region, most likely as a result of alternative splicing (Fig. 4). The significance of the presence of alternative exons within the 5′ untranslated region, all expressed in the same cell line, is unclear. Additional alternatively spliced p54$^{nrb}$ mRNAs may exist, since we only analyzed five out of dozens of positive clones. In all three cDNA classes, the available sequence upstream of the first Met codon contains in-frame nonsense codons. Therefore, this shared Met codon, which is present within an acceptable consensus initiation sequence (31), is presumed to be the start codon. The three classes of cDNA clones all end at the same position as the one shown in Fig. 3, 75 nt downstream of the common ochre codon. We did not find any polyadenylation

```
Clone 1                                       ggccgtgtagcgtcgccgttactccgaggaga    32
Clone 2,3,5                                    gccgtgtagcgtcgccgttactccgaggaga    31
Clone 4          gctcttttctcgggacgggagaggccgtgtagcgtcgccgttactccgaggaga    54

Clone 1          taccagtcggtagagg--------------------------------------    48
Clone 2,3,5      taccagtcggtagaggaggcccagtatttaggcgacagtgaatttattactctg    85
Clone 4          taccagtcggtagaggagaagtcgaggttagagggaactgggaggcactttgct   108

Clone 1          ------------------------------------------------------    48
Clone 2,3,5      aagagggttctgcacatatttccaaattatattggtggtcatcagaataggtga   139
Clone 4          gtctgcaatcgaagttgag-----------------------------------   127

Clone 1          ---------------------ggtgcaaaaATG                         57
Clone 2,3,5      taggaagaaatacttctcaagggtgcaaaaATG                        169
Clone 4          ---------------------ggtgcaaaaATG                        136
```

**Figure 4.** Comparison of 5′ untranslated regions of five p54[nrb] HeLa cDNA clones. Nucleotide residues are numbered on the right. The sequences of clones 2, 3 and 5 are identical. All five cDNA sequences are immediately preceded by the *Eco*R I linker GGAATTCC. The simplest interpretation of these structures is that the 5′ end of the p54[nrb] gene comprises at least four exons, with the initiation codon (bold uppercase) starting 10 nt into exon 4. Clones 2, 3, and 5 skip alternative exon 3 (57 nt), whereas cDNA clone 4 skips alternative exon 2 (114 nt), and cDNA clone 1 skips both alternative exons 2 and 3. Presumptive alternative exons 2 and 3 (bold lowercase) appear to be mutually exclusive; their order in genomic DNA cannot be deduced from our data. The first 22 or 23 nt of presumptive exon 1 in cDNA clone 4 are missing from the remaining clones, perhaps due to incomplete first strand cDNA synthesis or to nibbling of the DNA in steps prior to ligation of the *Eco*R I linkers.

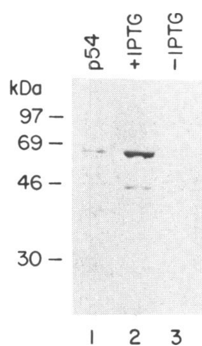


**Figure 5.** Recombinant p54[nrb] reacts with anti-PRP18 antibodies and has the same electrophoretic mobility as purified HeLa p54[nrb]. Protein samples were analyzed by Western blot as in Fig. 1. Lane 1: control gel-purified HeLa p54[nrb]; lane 2: 0.2 μl of induced *E.coli* cells harboring the p54[nrb] expression plasmid; lane 3: 0.2 μl of mock-induced cells. The fast migrating band seen only with induced cells (lane 2) is probably a degradation product of p54[nrb].

signals in the 3′ untranslated region. It appears that all the clones were generated by cleavage at an internal *Eco*R I site during construction of the original library. Since the stop codon was the same for all three cDNA classes, we did not attempt to obtain longer clones to sequence the complete 3′ untranslated region.

The p54[nrb] ORF encodes a 471 amino acid polypeptide with a calculated molecular weight of 54,267, which is about 10% smaller than the apparent molecular weight of the protein purified from HeLa cells, as determined by SDS-PAGE (Fig. 2). The three peptide sequences derived from purified HeLa p54[nrb] (see above) are present in the open reading frame of the cloned cDNA (Fig. 3). The p54[nrb] sequence bears no homology to yeast PRP18, although a few very short peptides might be shared epitopes responsible for cross-reactivity between p54[nrb] and anti-PRP18 antibodies. For example, the peptides RWKA and RSRL are present in p54[nrb], compared to RWKA and RTRL in PRP18 (14). Analysis of the deduced p54[nrb] protein sequence with the PC/Gene program Prosite (32) showed the presence of three

```
P54   54  GLTIDLKNFRKPGEKTFTQRSRLFVGNLPPDITEEEMRKLFEKYGKAGEVFIHKDK  109
PSF  277  GFKANLSLLRRPGEKTYTQRCRLFVGNLPADITEDEFKRLFAKYGEPGEVFINKGK  332
BJ6  282  GPTFELEPVEVPTETKPLGPNRLFVGNLTNDITDEELREMFKPYGEISEIFSNLDK  337

P54  110  GFGFIRLETRTLAEIAKVELDNMPLRGKQLRVRFACHSASLTVRNLPQYVSNELLE  165
PSF  333  GFGFIKLESRALAEIAKAELDDTPMRGRQLRVRFATHAAALSVRNLSPYVSNELLE  388
BJ6  338  NETEKKVNYIPNREKAARALEGRMRGGRQLRVRFAPNATILKVSNLSPFVSNELLY  393

P54  166  EAFSVFGQVERAVVIVDDRGRPSGKGIVEFSGKPAARKALDRCSEGSFLLTTFPRP  221
PSF  389  EAFSQFGPIERAVVIVDDRGRSTGKGIVEFASKPAARKAFERCSEGVFLLTTTPRP  444
BJ6  394  KSFEIFGPIERASITVDDRGHMGEPGIVEFAKKSSASACQRMSNEKCFFLTAALRP  449

P54  222  VTVEPMDQLDDEEGLPEK-LVIKNQQFHKEREQPPRFAQPGSFEYEYAMRWKALIE  277
PSF  445  VIVEPLEQLDDEDGLPEK-LAQKNPMYQKERETPPRFAQHGTFEYEYSQRWKSLDE  500
BJ6  450  CVVEPVEVNDETGGLPEKAFNKKMNDRQQEQSIGPPRADMNRFEHEYGSRWKQLHN  505

P54  278  MEKQQQDQVDRNIKEAREKLEMEMEAARHEHQVMLMRQDLMRRQEELRRMEELHNQ  333
PSF  501  MEKQQREQVEKNMKDAKDKLESEMEDAYHEHQANLLRQDLMRRQEELRRMEELHNQ  556
BJ6  506  LFNTQQDALKREQEMEEQKLEAQMLYKRYEQETEMLRQQLRQHEVRNEEKKLEWEM  561

P54  334  EVQKRKQLELRQEEERRRREEEMRRQQEEMMRRQQEGFKGT                374
PSF  557  EMQKRKEMQLRQEEERRRREEEMMIRQREMEEQMRRQREES                597
BJ6  562  REKQAEENRKREEETMRRQQTEMQSQMNRQEEDMLRRQQET                602
```

**Figure 6.** Sequence homology between central domains of human p54[nrb], human PSF and *Drosophila* NONA/BJ6. Identical residues are indicated by black background; conservative substitutions are highlighted in gray. Amino acid residues in each protein are numbered on both sides.

overlapping potential bipartite nuclear targeting sequences (33) at positions 324–340, 336–352, and 337–353. The presence of these sequences is consistent with the subcellular fractionation data (Fig. 1B).

To ensure that we had cloned cDNAs encoding the PRP18 cross-reacting protein, and that we had obtained the complete p54[nrb] coding sequence, we subcloned the p54[nrb] ORF into the *E.coli* expression plasmid vector pET-9c (26). After inducing the expression of recombinant p54[nrb] protein with IPTG, bacterial lysates were analyzed by Western blot with anti-PRP18 antibodies (Fig. 5). Recombinant p54[nrb] had identical electrophoretic mobility to p54[nrb] purified from HeLa cells, and retained immunoreactivity with anti-PRP18 antibodies.

## Sequence homology between p54[nrb], PSF, and NONA/BJ6

Comparison of the p54[nrb] protein sequence with those of HUM241D5 and *Drosophila* NONA/BJ6 revealed a high degree of homology. The HUM241D5 sequence was originally reported as a cDNA fragment encoding a human myoblast cell surface antigen (28). However, recent work showed that HUM241D5 represents the C-terminal part of PSF, a recently identified HeLa cell splicing factor of 712 aa (17). Evidently, the monoclonal antibody 241D5, which stains myoblast cell surfaces (28), fortuitously cross-reacted with PSF upon screening of a myoblast expression library (17). The 241D5 monoclonal antibody did not recognize human p54[nrb] on Western blots (data not shown).

A sequence alignment of human p54[nrb], human PSF, and *Drosophila* NONA/BJ6 is shown in Fig. 6. No gaps were introduced to generate the alignment, except for an extra Ala in NONA/BJ6. Human p54[nrb] and human PSF are 70% identical and 7% similar in an internal region of 320 aa. This region contains two consensus RNA recognition motifs (RRMs), a feature found in many RNA-binding proteins (reviewed in 18). *Drosophila* NONA/BJ6, a protein of 700 aa (29,30), has a 321 aa region with 42% identity and 7% similarity to the same 320 aa region of p54[nrb], and 44% identity and 9% similarity to that of PSF (Fig. 6). In addition, p54[nrb] and PSF are similar in their N-terminal regions in that both are rich in Pro and Gln residues (Fig. 3; ref. 17). The N-terminus of NONA/BJ6 is rich in Gly, Asn and Gln residues (29,30). All three proteins are also highly basic and their SDS-PAGE migration is slower than expected on the basis of calculated molecular weight. The homology between NONA/BJ6 and PSF was not noted previously, although the RRMs in both proteins were reported (17,29,34). We note

that the two RRMs comprise only 153 aa of the 320 aa homology region, and that the similarity between otherwise unrelated RRMs ranges from 10% to 18% identity (35).

### Northern and Southern blot analysis of p54$^{nrb}$

Since 320 out of 471 aa in p54$^{nrb}$ are in a region homologous to PSF, and in addition both proteins have N-termini with long stretches of Pro and Gln residues, we used a *Sac* I restriction fragment corresponding to the unique C-terminus of p54$^{nrb}$ as a probe for Northern and Southern blot analysis, in order to avoid cross-hybridization with PSF. Northern blot analysis (Fig. 7) of HeLa cell polyadenylated RNA at high stringency showed an abundant mRNA of approximately 2.6 kb and a rare mRNA of approximately 1.9 kb. The lengths of the five cDNAs range from 1,545 to 1,657 bp (Figs. 3 and 4). Assuming an average poly(A) length of 200 nt, the cDNA clones appear to be missing about 800 nt of untranslated sequence, most of which is probably from the 3′ end, since all the cDNAs were truncated upstream of their presumptive polyadenylation sites.

Under stringent Southern blot washing conditions (Fig. 8), the probe hybridizes to a single *Nco* I fragment of human genomic DNA, but to two or more fragments when the same DNA is digested by either *Eco*R I, *Hind* III, or *Pst* I. The probe encompasses a cDNA fragment from the *Sac* I site at +1021 to the 3′ end at +1491 (Fig. 3). The cDNAs have an *Nco* I site at +677 but lack internal sites for the remaining three enzymes. A single 3 kb *Nco* I genomic DNA fragment hybridized to the probe (Fig. 8, lane 3); the second genomic *Nco* I site must be located within an intron or downstream of the coding region. Hence, the fact that a single *Nco* I band was detected suggests that p54$^{nrb}$ is encoded by a single-copy gene. Since two or more bands hybridize to the same probe when genomic DNAs are digested by the remaining enzymes, which do not cut within the cDNA portion of the probe, we conclude that one or more introns interrupt the p54$^{nrb}$ gene in the 3′ terminal region encompassed by the cDNA fragment probe.

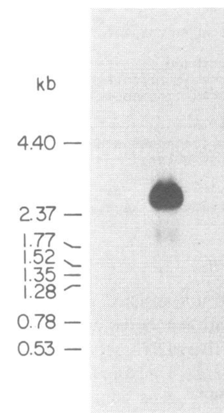


**Figure 7.** Analysis of p54$^{nrb}$ mRNA from HeLa cells. Polyadenylated RNA from HeLa cells was analyzed by Northern blot with a p54$^{nrb}$-specific cDNA probe fragment, as described in Materials and Methods. The relative mobilities of RNA size markers (BRL) are shown on the left. An abundant transcript of approximately 2.6 kb and a minor one of 1.9 kb are observed.

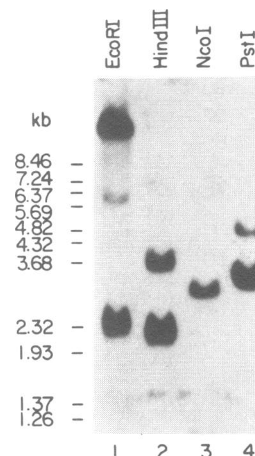


**Figure 8.** Analysis of the human p54$^{nrb}$ gene. Human genomic DNA was analyzed by Southern blot with a p54$^{nrb}$-specific probe. 20 μg of human genomic DNA was digested with *EcoR* I (lane 1), *Hind* III (lane 2), *Nco* I (lane 3), or *Pst* I (lane 4), fractionated by agarose gel electrophoresis, blotted onto a charged nylon membrane, and probed with a p54$^{nrb}$-specific cDNA probe fragment, as described in Materials and Methods. The relative mobilities of *Bst*E II-digested λ DNA size markers are shown on the left.

## DISCUSSION

A polypeptide from HeLa cells reacted very strongly in immunoblots with antibodies raised against the *S.cerevisiae* splicing factor PRP18. In addition, indirect immunofluorescence staining of HeLa cells with PRP18 antiserum produced a nucleoplasmic speckled pattern (D.Spector, unpublished data) typical of that generated by antibodies against snRNP particles and other splicing factors (16). These observations prompted us to look for a possible human homolog of this factor. Using reactivity with this antibody by Western blotting as a criterion, we purified p54$^{nrb}$ from HeLa cell nuclear extracts. Three classes of cDNAs coding for p54$^{nrb}$ were cloned based on partial peptide sequence of the purified protein. Sequence analysis of the cDNAs showed no significant homology between their common coding sequence and PRP18. We do not know why p54$^{nrb}$ reacts with anti-PRP18 antibody, although a few very short peptides present in both proteins might be common epitopes. At present, we cannot conclude whether there is a human homolog of yeast PRP18. Extensive screenings of human cDNA expression libraries have been carried out, either using anti-PRP18 antibody, or by attempting to complement the conditional growth phenotype of *S.cerevisiae prp18$^{ts}$* strains (unpublished data). No human homolog of yeast PRP18 has been found so far.

The translated sequence of p54$^{nrb}$ indicates that it is a basic protein with a calculated pI of 9.54, consistent with its chromatographic behavior, and a predicted molecular weight of 54,267, assuming no post-translational modifications. Its N-terminus contains runs of His, Pro, and Gln residues. The protein has two consecutive RNA recognition motifs, strongly suggesting that it is an RNA-binding protein. Preliminary UV crosslinking experiments have confirmed this prediction (data not shown). Further experiments are necessary to determine any sequence specificity in the binding of p54$^{nrb}$ to RNA.

Although the function of p54$^{nrb}$ is presently unknown, it has striking homology to the newly discovered HeLa cell splicing factor PSF (17) in a region that comprises 68% of the p54$^{nrb}$

residues. This region includes the two RRMs, as well as other sequences of unknown function. In addition, both PSF and $p54^{nrb}$ are unusually rich in Gln and Pro at their N-termini. These PQ domains are present in several hnRNP and snRNP polypeptides, and by analogy to similar domains in a number of transcription factors, they are thought to be involved in protein-protein interactions, and perhaps to have effector functions (17 and references therein). We speculate that $p54^{nrb}$ is also a splicing factor, and that PSF and $p54^{nrb}$ are different members of a family of splicing factors with related functions, but which may have, for example, different pre-mRNA substrate specificities. Additional experiments will be necessary to test whether $p54^{nrb}$ is required for pre-mRNA splicing *in vitro* and *in vivo*.

It is intriguing to find that the 320 aa region shared by $p54^{nrb}$ and PSF also exists in a *Drosophila* protein, NONA/BJ6. This protein is ubiquitously expressed in neural and non-neural nuclei throughout *Drosophila* development (34). It is one of three alternatively spliced isoforms encoded by the gene *no-on-transient A*, which is required for normal response to visual stimuli in *Drosophila* (29,30). *nonA* mutants show impaired phototaxis and optomotor responses, and their electroretinograms display aberrant on- and off-transient spikes, which reflect electrical currents that are thought to originate in the neurons of the lamina (30,34 and references therein). The *nonA* gene is also the locus of the *dissonance (diss)* allele, which causes aberrant male courtship song (34).

Although the molecular function of NONA/BJ6 is currently unknown and null alleles have not been isolated, the striking similarity to a splicing factor suggests that its pleiotropic functions in the central nervous system somehow involve pre-mRNA splicing. For example, NONA/BJ6 may affect alternative splicing of pre-mRNAs encoding ion channels or other proteins in such a way as to affect the expression of one or several isoforms that function in specific neural tissues.

The presence of NONA/BJ6 in transcriptionally active puffs on *Drosophila* polytene chromosomes (36) is consistent with the observation that splicing can take place on nascent transcripts (37,38). Interestingly, two other *Drosophila* puff-specific proteins are known to be splicing factors: B52 (39), which is a closely related variant of SRp55 (40,41), and RBP1 (42). Both proteins are members of the SR family of nuclear phosphoproteins, which appear to function in constitutive and alternative splicing (40–43). They are characterized by the presence of one or two RRMs and a C-terminal Arg- and Ser-rich domain. PSF, NONA/BJ6, and $p54^{nrb}$ have two RRMs but otherwise do not structurally resemble SR proteins. The subnuclear localization of PSF coincides with that of SR proteins and other splicing factors (16,17), and perhaps with that of $p54^{nrb}$, if the latter is responsible for the immunofluorescence pattern obtained with anti-PRP18 antibodies.

The extensive sequence homology in the 320 aa central domains of $p54^{nrb}$, PSF, and NONA/BJ6 strongly suggests conservation of function over the period of divergence of human and fruit fly, since the homologous segments of the three proteins align without any gaps, except for a single insertion of an Ala residue in NONA/BJ6. We propose to name this phylogenetically conserved region a DBHS domain (for *Drosophila* behavior and human splicing), based on the observed functions of its current representatives. It will be of interest to determine the functional consequences of swapping DBHS domains between $p54^{nrb}$, PSF, and NONA/BJ6 on splicing, visual acuity, and male courtship song. Since $p54^{nrb}$ and PSF are both expressed in HeLa cells, it is possible that these two proteins are representatives of a large family of DBHS domain proteins, which may function as general and/or regulatory splicing factors. Additional proteins with DBHS domains may be found, which would further show the range of size and sequence variation of DBHS domains and whether two RRMs are always included within the first half of the domain.

## ACKNOWLEDGEMENTS

We thank E.Birney and members of our lab for helpful discussions; J.Abelson for providing reagents prior to publication; F.Walsh for the 241D5 monoclonal antibody; J.Patton for communicating results prior to publication; D.Spector for immunofluorescence data; J.Posfai for PostScript software; A.Mayeda, D.Spector, and D.Helfman for comments on the manuscript. This work was supported by a grant from the National Institutes of Health. A.R.K. is a Pew Scholar in the Biomedical Sciences.

## REFERENCES

1. Moore,M.J., Query,C.C. and Sharp,P.A. (1993) In Gesteland,R.F. and Atkins,J.F. (eds.), The RNA World. Cold Spring Harbor Laboratory Press, New York, pp. 303–357.
2. Guthrie,C. and Patterson,B. (1988) *Annu. Rev. Genet.* **22**, 387–419.
3. Shuster,E.O. and Guthrie,C. (1990) *Nature* **345**, 270–273.
4. Pinto,A.L. and Steitz,J.A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8742–8746.
5. Anderson,G.J., Bach,M., Lührmann,R. and Beggs,J.D. (1989) *Nature* **342**, 819–821.
6. García-Blanco,M.A., Anderson,G.J., Beggs,J. and Sharp,P.A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3082–2086.
7. Smith,V. and Barrell,B.G. (1991) *EMBO J.* **10**, 2627–2634.
8. Liao,X.C., Tang,J. and Rosbash,M. (1993) *Genes Dev.* **7**, 419–428.
9. Rymond,B.C. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 848–852.
10. Birney,E., Kumar,S. and Krainer,A.R. (1992) *Nucleic Acids Res.* **20**, 4663.
11. Ruby,S.W. and Abelson,J. (1991) *Trends Genet.* **7**, 79–85.
12. Vijayraghavan,U., Company,M. and Abelson,J. (1989) *Genes Dev.* **3**, 1206–1216.
13. Vijayraghavan,U. and Abelson,J. (1990) *Mol. Cell. Biol.* **10**, 324–332.
14. Horowitz,D.S. and Abelson,J. (1993) *Mol. Cell. Biol.* **13**, 2959–2970.
15. Horowitz,D.S. and Abelson,J. (1993) *Genes Dev.* **7**, 320–329.
16. Huang,S. and Spector,D. (1992) *Current Biol.* **2**, 188–190.
17. Patton,J.G., Porro,E.B., Galceran,J., Tempst,P. and Nadal-Ginard,B. (1993) *Genes Dev.* **7**, 393–406.
18. Kenan,D.J., Query,C.C. and Keene,J.D. (1991) *Trends Biochem. Sci.* **16**, 214–220.
19. Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
20. Laemmli,U.K. (1970) *Nature* **227**, 680–685.
21. Hager,D.A. and Burgess,R.R. (1980) *Anal. Biochem.* **109**, 76–86.
22. Kawasaki,H., Emori,Y. and Suzuki,K. (1990) *Anal. Biochem.* **191**, 332–336.
23. Kawasaki,H. and Suzuki,K. (1990) *Anal. Biochem.* **186**, 264–268.
24. Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1989) In Current Protocols in Molecular Biology, J. Wiley & Sons, New York.
25. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) In Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
26. Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.* **185**, 60–89.
27. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.* **215**, 403–410.
28. Gower,H.J., Moore,S.E., Dickson,G., Elsom,V.L., Nayak,R. and Walsh,F.S. (1989) *Development* **105**, 723–731.
29. Besser,H.V., Schnabel,P., Wieland,C., Fritz,E., Stanewsky,R. and Saumweber,H. (1990) *Chromosoma* **100**, 37–47.

30. Jones,K.R. and Rubin,G.M. (1990) *Neuron* **4**, 711–723.
31. Kozak,M. (1989) *J. Cell Biol.* **108**, 229–241.
32. Bairoch,A. (1992) *Nucleic Acids Res.* **20**, 2013–2018.
33. Dingwall,C. and Laskey,R.A. (1991) *Trends Biochem. Sci.* **16**, 478–481.
34. Rendahl,K.G., Jones,K.R., Kulkarni,S.J., Bagully,S.H. and Hall,J.C. (1992) *J. Neuroscience* **12**, 390–407.
35. Birney,E., Kumar,S. and Krainer,A.R. Manuscript in preparation.
36. Frasch,M. and Saumweber,H. (1989) *Chromosoma* **97**, 272–281.
37. Sass,H. and Pederson,T. (1984) *J. Mol. Biol.* **180**, 9111–926.
38. Beyer,A.L. and Osheim,Y.N. (1988) *Genes Dev.* **2**, 754–765.
39. Champlin,D.T., Frasch,M. Saumweber,H. and Lis,J.T. (1991) *Genes Dev.* **5**, 1611–1621.
40. Mayeda,A., Zahler,A.M., Krainer,A.R. and Roth,M.B. (1992) *Proc. Natl. Acad. Sci.* **89**, 1301–1304.
41. Zahler,A.M., Lane,W.S., Stolk,J.A. and Roth,M.B. (1992) *Genes Dev.* **6**, 837–848.
42. Kim,Y-J., Zuo,P., Manley,J.L. and Baker,B.S. (1992) *Genes Dev.* **6**, 2569–2579.
43. Fu,X.-D., Mayeda,A., Maniatis,T. and Krainer,A.R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11224–11228.